数理解析研究所講究録
666巻 1988年 235-244

非正規関係集合の構造に対応するネットワーク構造

九州大学工学部 古川哲也（Tetsuya Furukawa）
九州大学工学部 上林彌彦（Yahiko Kambayashi）

1．まえがき

非正規関係は，冗長性が少なく，利用者に分かり易いデータベースの利用者インタフェースとして優れている[1][2]．これまでの非正規関係に関する研究は，データが満たす多値従属性集合などの意味制約と非正規構造との対応を明確にし，与えられた意味制約を満足するデータの非正規構造による表現[3][4]，最も冗長性の少ない非正規構造の設計[5][6]，非正規関係上での演算[7]等に関するもであった．しかし，利用者インタフェースとしての非正規関係を考えると，データベースにおける質問処理の結果を与えられた非正規構造でいかに効率よく出力するかが重要となる．

データベースシステムは，データのモデル化の方法により，関係データベースシステムとネットワークデータベースシステムに大別できる．効率の良さから広く普及しているネットワークデータベースは，そのリンク構造が非正規関係の構造と直接対応しており[8]，利用者ビューを非正規関係とするデータベースの実現に適している．本稿では，ネットワーク構造で表現できる非正規関係集合の構造を明確にし，ネットワークデータベースで出力できる非正規関係について議論する．

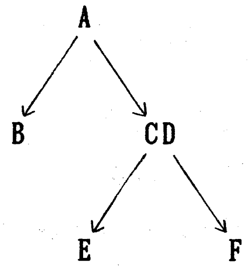
2．ネットワーク構造で直接表現される非正規構造

属性集合X上の関係R(X)は，属性値の組の集合である．各属性値が単純値のとき関係Rを正規関係，そうでないとき非正規関係という．関係を非正規化する代表的な操作としてRow-nest操作とGroup-by操作がある．関係R(X)に属性集合YによるRow-nest操作を適用したものは，X-Yが同じ値になる組集合をYの値を集合値にして１つの組としたものである．YによるGroup-by操作は，X-YによるRow-nest操作

として定義される．非正規構造は次のような有向木T(V,E)で表すことができる．Vは属性集合に対応する節点の集合である．節点v_1, v_2（$\in V$）で$v_1 \neq v_2$であれば$v_1 \cap v_2 = \phi$である．節点vを根とするTの部分木をT(v)，T(v)に含まれる属性の集合を$U_{T(v)}$，vから根までの属性の集合をA(v)とする．枝$<v_1, v_2>$は，$U_{T(v_1)}$上の部分関係に$U_{T(v_2)}$でのRow-nest操作を適用したことを表している．図１は(a)の正規関係を(b)の非正規構造に従って非正規化した関係(c)を示している．

A	B	C	D	E	F
a_1	b_1	c_1	d_1	e_1	f_1
a_1	b_1	c_2	d_1	e_2	f_2
a_1	b_2	c_1	d_1	e_1	f_1
a_1	b_2	c_2	d_1	e_2	f_2
a_2	b_3	c_3	d_2	e_2	f_3
a_2	b_3	c_3	d_2	e_2	f_4
a_2	b_4	c_3	d_2	e_2	f_3
a_2	b_4	c_3	d_2	e_2	f_4

(a)



(b)

A	B	C	D	E	F
a_1	b_1, b_2	c_1	d_1	e_1	f_1
		c_2	d_1	e_2	f_2
a_2	b_3, b_4	c_3	d_2	e_2	f_3, f_4

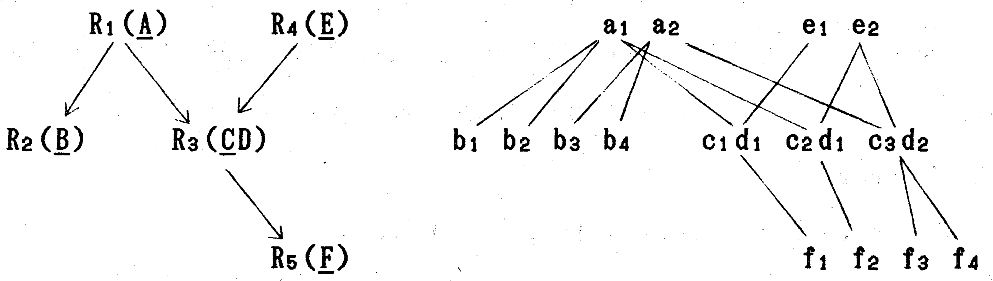
(c)

図１　非正規構造と非正規関係

関係R(U)で，属性集合Xの値を定めると属性集合Yの値の集合がU-X-Yの値に独立に定まるとき，Rは多値従属性$X \rightarrow\rightarrow Y$を満足しているという．$X \rightarrow\rightarrow Y_1, X \rightarrow\rightarrow Y_2, \cdots, X \rightarrow\rightarrow Y_n$のとき$X \rightarrow\rightarrow Y_1 | Y_2 | \cdots | Y_n$とかく．非正規構造Tでは，各節点v，vの子節点$v_1, v_2, \cdots, v_n$に対し，A(v)の値を定めると$U_{T(v_i)}$の値の集合は独立に定まるので，多値従属性$A(v) \rightarrow\rightarrow U_{T(v_1)} | U_{T(v_2)} | \cdots | U_{T(v_n)}$を表している．図１(b)の非正規構造は$A \rightarrow\rightarrow B | CDEF, CD \rightarrow\rightarrow E | F$を表している．属性集合Xの値を定めると属性集合Yの値が一意に定まるとき，関数従属性$X \rightarrow Y$を満足しているという．非正規構造Tで，節点vに対し，$v \rightarrow A(v)$であればvの値は重複しない．

ネットワークモデルは，同じ属性からなるレコードの集合であるレコード型と２つのレコード型間のレコードの１対多の対応を表す親子集合型の集合によって表現される．属性集合XからなるレコードR型をR(X)で，親レコード型がR_O,子レコード型がR_Mである親子集合型を$<R_O, R_M>$で表す．この構造はバックマン線図と呼ばれる有向グラフB(V,E)で表される．ここで，Vは各レコード型に対応する節点集合，Eは親レコード型から子レコード型へ向かう有向枝の集合である．値が定まればレコード型R(X)のレコードがただ１つ定まるような最小の属性集合KをRのキー

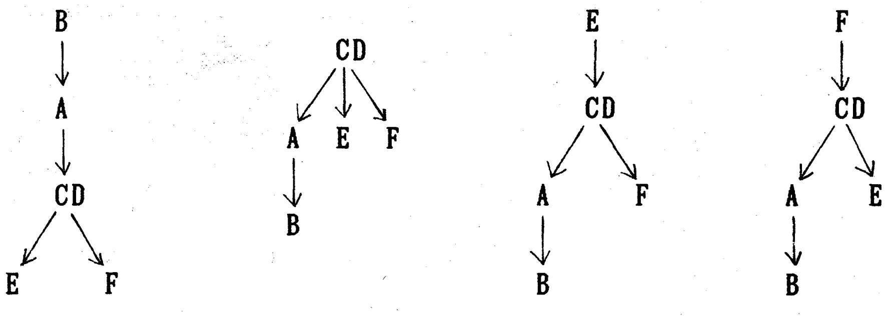
といい，K(R)で表す．ネットワーク構造では，各レコード型で，キーの値を定めるとレコードが定まり，レコードを定めるとその祖先のレコードが定まるため，関数従属性を表している．また，レコード型を除いたときの連結成分では，レコードの対応は独立なので，多値従属性を表している．図２は，バックマン線図とその実現値の例である．このバックマン線図は，関数従属性集合{$B \to A, C \to ADE, F \to ACDE$}，多値従属性集合{$A \to\to B|CDEF, CD \to\to AB|E|F$}を表している．



(a)　　　　　　(b)

図２　バックマン線図とその実現値

木構造のバックマン線図Bで，根レコード型Rがn個の子$R_1, R_2, \cdots, R_n$（木構造での子で親子集合型での子でなくともよい）持っているとする．R_iを根とする部分木のレコードの対応は，Rのレコードごとにまとまっており，それらはRの各レコードに対して直積的に対応する．これは，RによるGroup-by操作，R_iを根とする部分木によるRow-nest操作に対応している．図２(a)のバックマン線図で，R_1を根とすると，その非正規構造は図１(b)で表され，図２(b)の実現値は図１(c)の非正規関係となる．レコード型R_2, R_3, R_4, R_5を根とするとそれぞれ図３(a)～(d)の非正規構造となる．



(a)　　　　(b)　　　　(c)　　　　(d)

図３　ネットワーク構造が直接表現する非正規構造

［定義１］ バックマン線図Ｂで表されるネットワーク構造で，Ｂ中のレコード型を１つ定め，それを根とした有向木となるように各枝の方向を変え，各節点のレコード型R(X)を属性集合Xとした有向木で表される非正規構造を，Ｂが直接表現する非正規構造といい，そのような非正規構造の集合をUR(Ｂ)で表す．□

３．コストのかからない非正規構造の変換

ネットワーク構造は，非正規構造を直接表現していなくても等価的に表現している場合がある．非正規関係を他の構造の非正規関係に変換するときに，特別な操作を必要としない場合で，その様な変換には，次の３つがある．

(1) 入れ子の展開

(1-1) Group-by操作でまとめた値を展開する．

(1-2) Row-nest操作で直積となった値の集合を展開する．

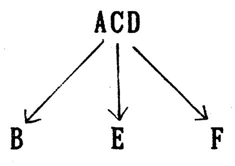
図１(c)の非正規関係の属性AによるCDのGroup-by操作を展開すると，図４(a)となる．その構造は図４(b)で表される．さらにACDによるBのGroup-by操作を展開す

A	C	D	B	E	F
a_1	c_1	d_1	$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$	$\begin{bmatrix} e_1 \\ \ \end{bmatrix}$	$\begin{bmatrix} f_1 \\ \ \end{bmatrix}$
a_1	c_2	d_1	$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$	$\begin{bmatrix} e_2 \\ \ \end{bmatrix}$	$\begin{bmatrix} f_2 \\ \ \end{bmatrix}$
a_2	c_3	d_2	$\begin{bmatrix} b_3 \\ b_4 \end{bmatrix}$	$\begin{bmatrix} e_2 \\ \ \end{bmatrix}$	$\begin{bmatrix} f_3 \\ f_4 \end{bmatrix}$

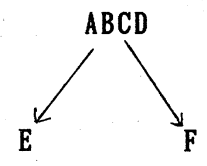
(a)

A	B	C	D	E	F
a_1	b_1	c_1	d_1	$[e_1]$	$[f_1]$
a_1	b_1	c_2	d_1	$[e_2]$	$[f_2]$
a_1	b_2	c_1	d_1	$[e_1]$	$[f_1]$
a_1	b_2	c_2	d_1	$[e_2]$	$[f_2]$
a_2	b_3	c_3	d_2	$\begin{bmatrix} e_2 \\ \ \end{bmatrix}$	$\begin{bmatrix} f_3 \\ f_4 \end{bmatrix}$
a_2	b_4	c_3	d_2	$\begin{bmatrix} e_2 \\ \ \end{bmatrix}$	$\begin{bmatrix} f_3 \\ f_4 \end{bmatrix}$

(c)



(b)



(d)

図４　入れ子の展開

ると図４(c)となり，その構造は図４(d)である．図４(c)は，図１(c)のAからの枝分かれによるRow-nest操作を展開したものになっている．

(2) Group-by操作の変更

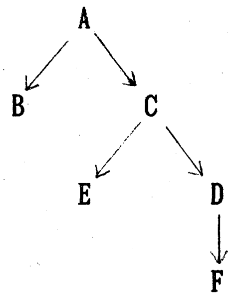
(2-1) Group-by操作の分離

(2-2) Group-by操作の順序の入れ換え

図５(a),(c)は，図１(c)の非正規関係のCDによるGroup-by操作をCおよびDのGroup-by操作に分離したものである．非正規構造はそれぞれ図５(b),(d)となる．図５(a)の非正規構造のDおよびFの順序を入れ換えたものが図５(e)であり，その構造は図５(f)である．

A	B	C	E	D	F
a_1	b_1	c_1	e_1	d_1	f_1
	b_2	c_2	e_2	d_1	f_2
a_2	b_3	c_3	e_2	d_2	f_3
	b_4				f_4

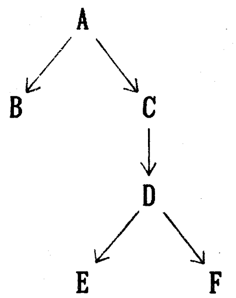
(a)



(b)

A	B	C	D	E	F
a_1	b_1	c_1	d_1	e_1	f_1
	b_2	c_2	d_1	e_2	f_2
a_2	b_3	c_3	d_2	e_2	f_3
	b_4				f_4

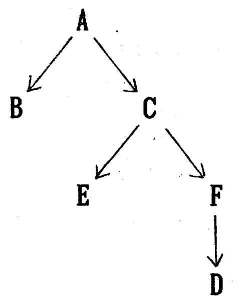
(c)



(d)

A	B	C	E	F	D
a_1	b_1	c_1	e_1	f_1	d_1
	b_2	c_2	e_2	f_2	d_1
a_2	b_3	c_3	e_2	f_3	d_2
	b_4			f_4	d_2

(e)



(f)

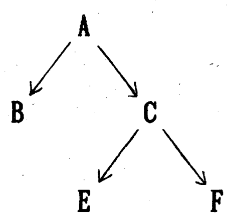
図５　Group-by操作の変更

(3) 属性の削除

図１(c)の非正規関係から属性集合CDを除くと，図６(a)の非正規関係が得られる．この構造は図６(b)である．

A	B	C	E	F
a_1	b_1, b_2	c_1, c_2	e_1 / e_2	f_1 / f_2
a_2	b_3, b_4	c_3	e_2	f_3, f_4

(a)



(b)

図６　属性の削除

4．ネットワーク構造で表現できる非正規構造

３節での非正規関係の変換に対応する非正規構造の変換は次のものである．

［変換１］　非正規構造Ｔ(V,E)を節点$v_1, v_2 \in V$，枝$(v_1, v_2) \in E$について，次のような非正規構造Ｔ'(V',E')にする．

$V' = V - \{v_1, v_2\} \cup \{v (= v_1 v_2)\}$,

$E' = E - \{(v_1, v_2)\} - \{(v_0, v_1)\} - \{\cup_i (v_1, v_{1i})\} - \{\cup_i (v_2, v_{2i})\}$

$\qquad \cup \{(v_0, v)\} \cup \{\cup_i (v, v_{1i})\} \cup \{\cup_i (v, v_{2i})\}$

ここで，v_{ki}はv_kの子節点，v_0はv_1の親節点である．□

［変換２］　非正規構造Ｔ(V,E)を節点$v, v_i \in V$，枝$\{(v, v_i)\} \subseteq E$について，次の(1)または(2)の非正規構造Ｔ'(V',E')にする．

(1) $V' = \{v$を$v' (v' \subset v)$で置き換え，$v'' (= v - v')$を加える$\}$

$E' = E - \{(v', v_i)\} \cup \{(v', v'')\} \cup \{(v'', v_i) \mid (v_i$はvの子節点$)\}$

(2) $v (= v'')$と$v' (\in \{v_i \mid v$の子節点$\})$を入れ換える．□

［変換３］　非正規構造Ｔ(V,E)，属性Aについて，Aを含む節点vを$v' (= v - A)$で置き換える．$v' = \phi$となるときはvを除き，vの親v_0と子$v_{M1}, v_{M2}, \cdots, v_{Mn}$に対し，枝

$(v_0, v_{Mi})(1 \leqq i \leqq n)$を加える．□

以下，非正規関係の変換を構造の変換で議論する．

［定義２］ 非正規構造Ｔが非正規構造Ｔ'に対する変換１～３の適用により得られるとき，Ｔ＜Ｔ'，特に，変換ｉのみで得られるとき，$T <_i T'$で表す．□

変換１はいかなる非正規構造に対しても適用できるが，変換２，３は常に適用できるわけではない．

［性質１］ 変換２は，$A(v)\text{-}v'' \rightarrow v''$のときのみ適用可能である．このとき，変換後の根からv'までの経路の値の組に対し，v''の値は一意なので何の操作も必要としない．この関数従属性が成り立たないとき，v''の値は複数存在し得るので，$v'v''$の部分関係でv'によるGroup-by操作が必要となる．□

［性質２］ 変換３は，$A \neq v$のとき$A(v)\text{-}A \rightarrow A$，$A=v$のとき$A(v_i)\text{-}A \rightarrow A$($v_i$はvの子節点)のときのみ適用可能である．これらの条件を満たさないとき，$A \neq v$のときは$A(v)\text{-}A$の値の組に，$A=v$のときは$A(v_i)\text{-}A$の値の組に重複が存在し得るので，Group-by操作が必要となる．□

［定義３］ 非正規構造Ｔに対し，Ｔ'＜Ｔとなる非正規構造をＴで表現できる構造といい，そのような構造の集合をRep(Ｔ)で表す．非正規構造の集合$\mathbf{T}$に対しては，$\mathrm{Rep}(\mathbf{T})=\{T' \mid T' \in \mathrm{Rep}(T), T \in \mathbf{T}\}$とする．□

定義３より，与えられた非正規構造の集合$\mathbf{T}$を表現するネットワーク構造を議論する場合，$\mathbf{T} \in \mathrm{Rep}(\mathbf{T}')$となる非正規構造の集合$\mathbf{T}'$を考えればよい．従って，非正規構造の集合を表現するネットワーク構造の設計では，次の点に注意すればよい．

(1) なるべく非正規化されたもの，即ち非正規構造Ｔに対し，$T <_1 T'$となる非正規構造Ｔ'を用いる（変換１）．

(2) Group-by操作については，非正規化されていなくてもよい場合がある（変換2，性質1）．
(3) 属性を多く用いる（変換3）．

(2)は次の既約化された非正規構造で代表した方がよいことを意味する．

［定義4］ 有向木T(V,E)で表される非正構造は，根v_0以外の節点vで，根からvまでの経路$v_0, v_1, \cdots, v_n, v$で$\cup v_i \rightarrow A (A \in v)$となるものが存在しないとき既約であるという．既約でないとき，vを$v' = v - A$とし，$\cup_{j=0,k} v_j \rightarrow A$となる最小のkの$v_k$にAを加え既約にする操作を既約化という．□

既約化した非正規関係が表現可能であれば，各組のv_kのAの値をT(v')の集合値のAの値とすることにより元の非正規関係が得られる．これは，$\cup v_i \rightarrow A$より$\cup v_i$が定まればAが定まるので，T(v)でRow-nest操作を行うとT(v)の値の集合のAの値が一致するためである．

［定理1］ 非正規構造T，およびTを既約化した非正規構造T'は，$T <_2 T'$であり，$T'' \neq T'$，$T <_2 T''$，$T' <_2 T''$となる非正規構造T"は存在しない．□

（証明） 既約化を行う操作の逆の操作は変換2で実現でき，関数従属性が成り立つので性質1より適用可能で，$T <_2 T'$となる．$T' <_2 T''$となるT"が存在すれば，T'は既約でなく，既約化によりT"とすることができる．（証明終り）

5．オブジェクトによる非正規構造の表現可能性

意味的にまとまりのある属性集合をオブジェクトという．本節では，ネットワーク構造のオブジェクトと非正規構造のオブジェクトを比較することにより非正規関係を表現可能であるかどうかを判定する．

［定義5］ バックマン線図Bで表現されるオブジェクトZは，属性集合Zを含むBの部分ネットワーク構造B'で，すべてのレコードの組のZの値の組に重複がな

いようなＢ’が存在する属性集合Zである．□

Ｂの連結な部分グラフＢ’で表されるネットワーク構造（Ｂの部分ネットワーク構造）に含まれる属性集合$U_{B'}$も表現可能なオブジェクトであるが，データの対応を求める際にレコード型でキー以外の属性は無視することもできる．

［定理２］ ネットワーク構造Ｂで表現可能なオブジェクトは，Ｂの部分ネットワーク構造Ｂ’で，$\cup K(R_i) \subseteq Z \subseteq U_{B'}$（$R_i$はＢ’に含まれ，Ｂ’中に子レコード型を持たないレコード型）となるＢ’が存在する属性集合Zである．そのような属性集合Zの集合をOBJ(Ｂ)で表す．□

（証明） オブジェクトの定義より，$Z \subseteq U_{B'}$である．$\cup K(R_i) \subseteq Z$であれば，子レコードが定まれば親レコードが定まるので，$\cup K(R_i) \to U_{B'}$が成り立つことから，$\cup K(R_i)$の値に重複はない，即ちZの値の組に重複はない．$\cup K(R_i) \not\subseteq Z$であればZの値の組に重複が存在し得る．（証明終り）

非正規構造のオブジェクトを定義し，ネットワークデータベースでの非正規関係の表現可能性を示す．

［定義６］ 非正規関係の構造を表す有向木Ｔ(V,E)のオブジェクト集合を，OBJ(Ｔ)＝{Z｜Ｔの根から節点v(∈V)までの経路の属性集合}とする．□

［定理３］ 構造が有向木Ｔ(V,E)で表される非正規関係が，バックマン線図Ｂで表されるネットワークデータベースで表現可能であるための必要十分条件は，Ｔを既約化した構造Ｔ’がOBJ(Ｔ’)$\subseteq$OBJ(Ｂ)かつＴ’の表す多値従属性集合がＢの表す多値従属性集合に矛盾しないことである．□

（証明） Ｔ’がＢで表現可能であれば，Ｔ＜Ｔ’よりＴも表現可能である．このとき，根から各節点vまでの属性集合からなる部分関係には組の重複がないので，OBJ(Ｔ’)$\subseteq$OBJ(Ｂ)であり，ＢはＴ’の表す多値従属性集合Ｍに矛盾しない．Ｂが

Ｍを満足しなければ，そのようなデータはＴ’で表現できない．また，OBJ(Ｂ)に含まれないＴ’のオブジェクトが存在すれば，Ｂのそのオブジェクトの属性集合からなる部分に値の組の重複が存在し，Ｔ’の構造で表現できない．（証明終り）

６．むすび

ネットワーク構造で表現できる非正規構造の集合について議論した．非正規関係の集合を効率よく実現するときは，ネットワークモデルを用いる方がよい．そのためには，非正規構造の集合が与えられたときのそれらを表現するネットワーク構造の設計法を確立しなければならない．５節で示したオブジェクトはそのための道具として用いることができる[9]．

参考文献

1) Makinouchi,A., “A Consideration on Normalized Form of Not-Necessarily-Normalized Relation in the Relational Data Model,” Proc. Int. Conf. on Very Large Data Bases, pp.447-453, 1977.
2) Sholl,M.H. and Scheck,H.-J. (ed.), Theorey and Applications of Nested Relations and Complex Objects, Workshop Material, INRIA, April 1987.
3) 上林，田中，武田，矢島，“関係データベースにおける意味制約を反映した非正規関係の設計問題”，情報処理学会論文誌，第24巻，第6号，pp.877-885, 1983年.
4) Fisher,P.C., Saxton,L.V., Thomas,S.J., and Van Gucht,D., “Interactions between Dependencies and Nested Relational Structures,” J. Comp. and Syst. Sci., Vol.31, No.3, pp.343-354, Dec. 1985.
5) Ozsoyoglu,Z.M. and Yuan,L.-Y., “A New Normal Form for Nested Relations,” ACM Trans. Database Syst., Vol.12, No.1, pp.111-136, March 1987.
6) Takeda,K., “On the Uniqueness of Non-First-Normal-Form Relations,” Theorey and Applications of Nested Relations and Complex Objects, April 1987.
7) Jaeschke,G. and Scheck,H.-J., “Remarks on the Algebra of Non First Normal Form Relations,” Proc. ACM Symp. on Principles of Database Syst., pp.124-138, March 1982.
8) Kambayashi,Y. and Furukawa,T., “Unnormalized Relations for Network Database Output,” Theory and Applications of Nested Relations and Complex Objects, April 1987.
9) 古川，上林，“オブジェクト集合を用いたネットワークデータベースの設計”，情報処理学会データベースシステム研究会，昭和５３年７月.